STATISTICAL ASPECTS
OF CLINICAL TRIALS

BYRON WM. BROWN, JR.
STANFORD UNIVERSITY

1. Introduction

My object in this lecture will be to give an overview of the statistical aspects of current clinical trial methodology, including a very brief history, my view of current practice, recent relevant statistical developments, and areas that need further statistical research.

My primary concern will be clinical experiments, though I will have some comments on other types of study of clinical data also.

Of course, the idea of trying out a new treatment and then comparing the results with past experience with other remedies is natural. Insistence on some systematic attempt to assure a *controlled* comparison of treatment effects is a recent development. Scattered reports of controlled studies have appeared in the literature only within the last few hundred years. The idea of *random* allocation of treatments to experimental units, in agricultural science, originated with Fisher [30] and gained acceptance in agriculture through the work of such men as Yates and Snedecor [51]. My impression is that agricultural scientists now accept the ideas of randomization, experimental design, and statistical evaluation as essential to sure and orderly scientific progress.

These experimental principles were introduced into clinical medicine in the post World War II period by Hill [36], [37] and taken up by Mainland [43], Lasagna [41], and others. In recent years Cornfield [18] and Armitage [4] have played important roles in encouraging further development of the methodology for planning and evaluating clinical experiments.

2. Present practice

Clinical trial methodology, employing concurrent controls, randomization, and the blindfold technique, has had a great impact on medical scientists. Hundreds of valid medical experiments have been completed and currently it can safely be said that the method is accepted and in use by at least some investigators in every medical specialty. However, it must be quickly added that in many fields, such as surgery, randomized clinical trials are still rare. In fact, there are influential and fluent clinical scientists who enthusiastically point out the difficulties and publicly ponder the usefulness of the experimental approach in clinical

1

medicine. Some, indeed, are opposed to the whole idea of using the patient as an experimental subject.

Although randomized, even blindfold, trials have been used in many types of clinical investigation, the methodology has taken hold and been accepted as standard experimental technique only in the evaluation of drugs and immunizing agents. This is due in large part to the greater ease of application of the techniques in this field as contrasted with applications in surgery or radiation therapy, for example. In the United States, pressures from the Federal granting agencies, particularly the National Heart and Lung Institute, and from the Food and Drug Administration [2], as the regulatory body for approval of drugs for marketing, have hastened the use, if not the acceptance, of clinical trial methodology for the evaluation of drugs.

Even in clinical investigations of the efficacy of drugs the methodology is far from uniformly well accepted and practiced. The National Cancer Institute strongly encourages the use of randomized clinical trials in the evaluation of cancer treatments, yet in a recent meeting on the design of clinical studies in cancer, well known scientists raised their voices on each side of the issue. Chalmers [15] has recently reviewed the clinical cancer research literature and he states that only about 20 per cent of the clinical studies, reported in abstracts submitted in 1965–1970 to the American Association for Cancer Research, could be considered controlled experiments. Among this 20 per cent, only a portion would be randomized or double blind. Chalmers [14] has reported also that a recent survey of clinical trial abstracts submitted to an annual meeting of the American Gastroenterological Association revealed only 4.5 per cent contained evidence of adequate techniques to minimize investigator bias.

At the present time there are important clinical trials under way in areas of critical importance to the public health. A diet-heart study in Minnesota has most of the patients in the state mental institutions on double blind, randomly assigned normal or low fat diets; the purpose is to determine long term effects of low fat diet in preventing heart disease. The Coronary Drug Project [21] has thousands of men on randomly assigned drugs, including placebo, in order to discover which of the drugs are effective in preventing coronary heart disease. The UGDP (University Group Diabetes Program [40]), a study now in its 11th year of followup, has been a landmark in clinical research in diabetes. This study has shed new light on the use of insulin in diabetes therapy and has raised disturbing questions about the efficacy, and even the safety, of oral hypoglycemic substitutes for insulin in mild forms of diabetes.

Today analgesic agents and other psychoactive drugs used by the anesthetist and the psychiatrist are routinely evaluated in double blind, randomized experiments. The same is true of the study of antibiotics, and research in these fields is marked by sure and steady progress.

A new aspect of modern clinical research is the cooperative study, a study in which investigators from several hospitals or clinics follow the same study protocol and pool their results to obtain a definitive answer more quickly than any one

investigator could obtain alone with his limited supply of appropriate patients. These cooperative studies have become more and more common as medical investigators have begun to ask more subtle questions requiring larger sample sizes. The Coronary Drug Project [21], the Veterans Administration study of prostate cancer [55], and the University Group Diabetes Project [40] are good examples of current cooperative studies.

In any review of current clinical study methodology, it is important to mention the growing concern for information on the risk of side effects, especially those of relatively low incidence, associated with medical treatments. Sometimes these risks are uncovered in randomized trials carried out to evaluate efficacy, as in the VA and CDP trials (mentioned above) which disclosed that estrogen causes heart attacks. Most investigators feel that a randomized clinical trial, purposely designed to study an imputed side effect, would be unethical. This is one reason for the lack of experimental data on the carcinogenic effects of birth control pills, the risk of liver damage caused by the anesthetic agent, Halothane, and the various lethal effects attributed to smoking. Of course, a second very good reason for not studying low incidence side effects experimentally is the tremendous cost and effort involved in detecting the very small differences in rates involved.

3. Current statistical problems and recent work in the design of clinical trials

The objection most often raised to the randomized clinical trial is the ethical question of withholding from a patient by random choice what appears to be a new and better treatment, or, conversely, randomly assigning him to a new treatment, carrying unknown hazards, when he could be given the more familiar and reliable standard treatment. The most common rejoinder by clinicians and statisticians alike is that they would not randomize unless the competing treatments were preferred equally. To me, this has always seemed to be an evasion of a real issue. Taking all current information on the treatments into account, and taking into account specific information on the given patient, cases of absolutely equal preference for competing treatments would be rare. Indeed, this is why so many clinical scientists object so strenuously to the randomized trial. The issue must be met squarely in terms of prior probabilities, risks to the patient, counterbalancing benefits the patient derives from being in a carefully executed and generously staffed clinical experiment, and possibly the moral obligation of the patient to add a little information for the benefit of his fellow man when the added risk he might incur is small. Only when statisticians help to formalize this ethical dilemma in a simple and convincing way, for some real situations, only then, will we have helped the clinician and consulting statistician out of their dilemma, so that they can deal honestly with themselves, if not with the patients.

The organization and management of a clinical trial, especially a cooperative trial, are very important and there are some good guides on the design of a clinical trial. The writings of Hill [37], Mainland [43], and Feinstein [27], and

the books edited by Witts [56], and by Hill [35], would be extremely helpful to anyone who is designing a clinical study; the UGDP [40] and CDP [21] trials are models of organization and management for a cooperative trial.

It is always tempting to "use the patient as his own control," as the clinical scientist puts it. Oftentimes the patient can be given one treatment, later "crossed over" to the other treatment, and his response to therapy can be measured in each time period. This kind of design is used in allergy experiments, for example, and in other chronic diseases such as arthritis. Cochran and Cox [16], and D. R. Cox [22] discuss these designs. Most discussions assume the first treatment has no residual effects in the second period, but Grizzle [32] and others have discussed the difficulties of handling residual treatment effects. The design and analysis become complicated, however, and they often depend on the assumption of additive residual effects. The current tendency in practice seems to be to avoid the crossover where possible, even at the expense of necessarily larger sample size, because of the difficulties in interpreting crossover results.

In the case of completely randomized experiments, where the patients come to the physician and are admitted to the study one at a time, there is always a worry about lack of balance in numbers of patients assigned to the several treatments, as the trial subjects accumulate. This is even more important when the patients are blocked by type, and balance within each block is important. The method of pseudorandomization by alternating patients or admitting patients to one treatment or another on alternate days has long been discredited, after some unfortunate experiences. The usual method of assuring balance today is to randomize within groups of six or eight or ten patients arriving one after the other at the clinic, without disclosing the group size to the investigators who admit the patients to the study. Efron [26] has investigated this strategy and compared it to the completely randomized design and to what he calls the biased coin design. The latter design assigns the next treatment with an assignment probability that varies so as to tend to even the cases assigned to the several treatments at each step. Efron compares the designs with regard to susceptibility to several sources of bias, but he comes to no conclusions concerning the design of choice. Stigler [54] has also considered methods of eliminating or minimizing bias in randomized experiments.

Selection of patients is a question in any clinical trial. It is obvious that there is a distinct advantage in selecting patients who are moderately ill where this is possible, because severely ill patients won't respond to either treatment, and mildly ill patients will respond to any treatment. However, the clinical scientist, and especially the practicing physician, are leery of any selection of patients. Results are most credible to them when the experimental patients resemble their own patient population. This is partly due to an attitude developed in evaluating uncontrolled studies, where comparability with other sets of data was essential, and partly due to a valid concern for the degree to which the results of a randomized experiment can be generalized to other experimental units.

The specification of sample size for planning purposes, before a trial is started,

is a common problem to the biostatistician. Often he can use tables for the power of the two sample binomial test or two sample t test, as found in Owen [48], Schneiderman [50], Dixon and Massey [23], Snedecor and Cochran [52], or Mainland [43]. However, special problems arise in the planning of clinical trials. In particular, the endpoint or measurement made in the clinical trial is often a matter of *when* an event occurs (that is, a waiting time, perhaps to death) rather than *if* the event occurs. Since the patients are admitted to the study at varying times and the observations are terminated at a common time, namely, at the calendar time cutoff date for the study, the data present a problem in waiting time distributions, with censoring at a different time for each patient. Ederer [24] has presented a table that allows a computation of required number of patients, assuming an exponential waiting time distribution, a given patient admission rate, an estimate of the survival rate, and a specification of the desired standard error for the estimate of survival rate at a given time. Pasternack and Gilbert [49] have carried this work further and present some tables that are useful in planning.

A committee of the National Heart and Lung Institute [1] has presented tables of required sample size for the special clinical trials (for example, diet-heart trials) in which the treatment is not expected to take effect for some time after the commencement of therapy and there is a certain attrition on drop-out rate for patients during followup. Halperin, Rogot, Gurian, and Ederer [33] present the details of this work. A committee of the American Heart Association [11] has presented sample sizes for the same sort of clinical trial, ignoring the factors of delay time and dropout rate, and concentrating on the problem of the effect on required sample size of unreliability in the judgment of cause of death. Both of these papers are rather specific to the diet-heart question, though the results are of some general use.

Of course, the fact that patients are usually admitted to a clinical trial sequentially, and slowly enough to permit rather frequent, if not continuous, analysis of the data, suggests that sequential stopping rules for clinical trials would be most appropriate. The first to point this out seems to have been Bross [8], who published some truncated sequential plans for the one sample binomial case. He proposed that the patients be paired as they were admitted to the study, each pair being declared a win or loss for a given treatment, this Bernoulli variable furnishing the sequential data for the test. Armitage's book [4], contains a rather complete exposition of sequential analysis applied to clinical trials, and he presents plans for the case of a single Bernoulli variable, and also a normally distributed variable. Armitage also suggested a way of handling the analysis of exponential variables observed in survival time studies.

Miller [45] has recently published sequential plans for nonparametric sequential analysis, especially well suited to clinical trial application. The Miller plans are based on Monte Carlo sampling results and his work is concerned with the one sample case, so that pairing of patients is required. The procedure is based on the one sample signed rank statistic and the limits are set at a fixed multiple

of the conditional standard deviation of the statistic, so as to control the probability of going out of limits at any point from sample size one to a preassigned truncation point. Gehan [63] has developed a two sample procedure for variable censoring times, typical of clinical trial time data, based on a generalization of the Wilcoxon statistic. Efron [25] proposed modifications of the procedure that increase its power for certain alternatives. Breslow [7] has generalized the Gehan procedure to the case of more than two treatments under trial.

The statisticians for the UGDP (Meinert, Knatterud, and Canner) also used Monte Carlo sampling to develop a sequential plan for monitoring the diabetics in that clinical trial. However, they used as their waiting time distribution function the survival function defined by *U.S. Life Tables*. They assigned each patient in the trial the expected survival function appropriate for his age and sex at entry into the trial.

Anscombe [3], in his review of Armitage's book, was quite critical of the whole idea of looking on the clinical trial as a hypothesis testing situation. Anscombe's arguments are quite Bayesian in flavor and reminiscent of Fisher's criticism [29] of the Neyman-Pearson view of inference. Most statisticians today who are involved in the planning and evaluation of clinical trials would sympathize with Anscombe's views. Though they may use Neyman-Pearson theory to calculate a required trial size or to set up sequential limits for the trial, acceptance-rejection rules are not taken to be hard and fast. The data are continually scrutinized, analyzed in ways unanticipated at the start of the trial, and analyzed in the light of new information on additional ancillary variables, as the number of patients increases and makes such analyses possible.

Armitage himself was the first to suggest the adaptation of some work by Maurice [44] as a new and more practical way of calculating the required size of a clinical trial. The idea was to estimate the total number N of patients to be treated by one of two treatments, it being unknown which of the two treatments was better. A clinical trial with n patients per treatment was to be carried out, in order to decide on the better treatment, and the remaining $N - 2n$ patients would be treated with the chosen treatment. How large should the trial be, that is, what is the optimal choice of n? Colton [17] pursued the problem at the suggestion of Armitage, for the normal case, sigma known, with loss function equal to the difference in means for each use of the inferior treatment. He looked at the unknown difference between treatments from both a minimax and a Bayesian point of view. Colton obtained some interesting results, the most surprising of which was that the optimum clinical trial size in certain circumstances might be as much as $\frac{1}{3}$ of all patients to be treated, even when the total number of patients to be treated N was very large. Canner [13] pursued these ideas for the Bernoulli case, obtaining results similar to Colton's, and with additional results that demonstrated that the optimal sample size n did not depend very strongly on the total population size N. For example, in the case of uniform prior distributions Canner showed that the estimate of N could be off by a factor of two

without serious increase in the loss function. Canner also investigated the required trial size when the cost of clinical trial observations was added to the cost of treatment failures.

The results of Colton and Canner are really single stage solutions to the two arm bandit problem. Colton has published two stage results also. The underlying dilemma here is that of allowing a clinical trial to go on when the evidence is mounting in favor of one of the treatments. Cornfield, Halperin, and Greenhouse [20], Zelen [60], and Sobel and Weiss [53], have explored variations of a more radical solution, the play the winner strategy, where the next patient is assigned the treatment given the last patient if the last treatment was a "success"; otherwise, the alternate treatment is given. The play the winner strategy has the obvious advantage of tending to assign the better treatment to more of the patients and it does this more successfully than its competitors in a variety of circumstances. However, the strategy has practical difficulties for the clinical trial situation in that the outcome for the last patient is not often known at the time the next patient comes in; and more serious, the method of assignment, even if done on a random basis with varying probabilities causes difficulties in assuring the unbiasedness or blindfold aspect of the trial. A further difficulty that I have experienced is that the clinical scientist who can convince himself that random allocation with equal probabilities for the several treatments is ethical, cannot bring himself to allocate with unequal probabilities or to play the winner. In fact, when I suggested the strategy to some eminent clinical colleagues, they regarded it as a quite incredible proposal.

With regard to planning clinical trials, it should be mentioned that there are efforts among clinicians to work out the rationale for a trial design in terms of underlying mechanisms for the specific disease of interest. In leukemia, there are efforts to understand how the disease develops and how it reacts to chemotherapy and radiotherapy, and to choose therapy strategies accordingly. The clinical pharmacologist in general is concerned with the appropriate dose levels, times of administration and route of administration, with regard to the bio-availability of the drug with various strategies. The statisticians must also concern themselves with these facets of design; and, as this type of approach becomes more formal, statisticians find themselves acting as mathematical biologists, mapping out plans, even simulating the clinical trial itself. An example is Bross's mathematical modeling work [9] purporting to show that a proposed clinical trial in breast cancer would be futile.

4. Analysis of clinical trial data

Biostatisticians are familiar with the computer and usually have programming and computing resources available to them. The computer has produced an extraordinary revolution in the routine analyses that are done on data from clinical trials. I have already mentioned sequential analysis since it is so closely tied

to planning as well as analysis; most of the work of Armitage, Miller, Meinert, Knatterud, Canner, Zelen, and Sobel in deriving and evaluating sequential strategies depended to some extent on the use of computers.

Computers have made possible complex analyses of the typical waiting time data with varying censoring points so familiar in clinical trials. The actuarial methods like that of Berkson and Gage [5] using grouped data, can now be carried out easily without grouping, as suggested by Kaplan and Meier [38]. Furthermore, the analysis can be carried out each time a patient is seen again or a new death or other event is reported, by adding the new observation to the computer file, updating the file, and printing out the new survival curve, with confidence band. Gehan [31] has proposed methods for obtaining estimates of the hazard function or the force of mortality function and the density function, as well as the survival function, using actuarial methods, though these are not popular yet.

In the early fifties, Littell [42] proposed a maximum likelihood method for estimating an exponential survival function using typically censored clinical trial data. Now, with the aid of the computer, much more general parametric models are used and the estimates and standard errors are easily obtained. The generalizations go in several directions—the hazard function may be taken to be an appropriate function of time, linear or exponential, for example. Competing risk models may be used, with each risk a function of time. In particular, competing risks that are linear functions of time yield a convenient and useful model because the sum of the risks will also be linear in time.

Zelen and Feigl [61] and Zippin and Armitage [62] have furnished good examples of efforts to allow the hazard function to be dependent on ancillary variables, yet another avenue for the use of the computer that allows fuller exploration of the data. The work of Boag [6] and Berkson and Gage [64] and, more recently, the work of Haybittle [34] illustrate another approach to the formulation of parametric models for survivorship that are tailor made for specific diseases and treatment comparisons. The Berkson and Gage model, for example, allowed a cure rate following surgery, with the cured and noncured patients following different survival functions following surgery.

These efforts to extend classical significance testing and estimation to more satisfying models that allow for more realistic changes in the force of mortality, and the use of ancillary information, have been extremely important. However, more fundamental changes in methods of analysis have come about through the use of the computer in the evaluation of clinical trial data. First, along with the use of more complex models has come a heavier use of the likelihood function or likelihood contours, though such analyses are still not common in clinical journals. Second, permutation tests are coming into use. Third, Bayesian and semi-Bayesian procedures are coming into play, with Cornfield as a principal exponent [18], [19]. An example is the Cornfield approach [40] used in the UGDP report on the effects of Tolbutamide. Cornfield computes the ratio of the likelihood for the null hypothesis to the average likelihood over a set of alternatives,

where the averaging function is specified beforehand or else determined from the data by minimization. He calls these ratios relative betting odds (RBO's). Another example of Bayesian techniques applied to a set of clinical trial data can be found in a paper by Novick and Grizzle [47].

A fourth development or set of developments is typified by the reports on the Halothane Study [12]. This study was not a randomized clinical trial, but a comparison of the mortality results for persons getting Halothane as a general anesthetic at surgery *versus* those getting other agents. A number of statistical procedures were developed for adjusting for many ancillary variables as covariates, in comparing death rates. The procedures developed in the Halothane Study for handling large numbers of covariates in samples of modest size, are being applied to current randomized clinical trials.

5. Future statistical research and development

There are many areas in the statistical methodology for clinical trials that need the attention of the statistician.

5.1. Some specific examples of dilemmas in choice of scientific strategy should be studied formally, with the object of shedding some light on the question of when a randomized clinical *experiment* is indicated, as opposed to a retrospective or prospective study. The factors of cost, the possibilities of bias, the ethics and the question of scientific credibility should all be considered in a mathematical formulation of the problem.

5.2. Tables of sample size requirements should be generated that apply to the types of situation encountered in clinical trials, including varying forms of hazard function, patient accrual rates, and dropout rates.

5.3. The whole question of stopping rules should be considered with regard to the credibility of the final report. What does influence the scientific audience with regard to the way in which the decision to stop is made? What should influence them? How should the decision rule be reported?

5.4. The adjustment of results for a multitude of baseline variables, or covariates, must be considered. We need more methodological development and a closer look at the methods already developed such as those reported in the Halothane Study, but consideration of the interpretation of results is even more important. Can one test the validity of the randomization itself by checking the multitude of baseline variables? How should one adjust for carrying out a multitude of *a priori* and *a posteriori* inferences on the same set of data, and how should one report the results?

5.5. Methodology should be laid down for looking at the likelihood function for many parameters, and both statisticians and medical scientists must become practiced in looking at such presentations and interpreting them.

5.6. The clinical trial must not be regarded only as a tool for isolated experiments. It must be adapted to routine operations of the clinic or hospital, as an accepted part of normal practice. Kiresuk, Salasin and Sherman have reported

[39] that at the Hennepin County Mental Health Center in Minneapolis, every patient coming to the Center is given a standard workup. Objectives of therapy are set, followup plans are laid, and possible forms of treatment are listed (day clinic, group therapy, drugs, and so forth) by an Intake Committee. If several methods of treatment are thought to be feasible, the patient is randomly allocated to one of the competing treatments. Thus, clinical experimentation with systematic followup becomes a part of routine medical practice. Statisticians must work with the clinician to see that this kind of automatically evolving and improving system becomes the rule. Such a system will involve new concepts in clinical trial management, in statistical analysis, monitoring and decision making.

5.7.   These days we are concerned with health care systems and the quality of medical care. Again the statistician must consider adaptation of the clinical trial methodology to this area. If it is unethical to choose a treatment for one patient on the basis of unscientifically collected data, then it must be all the more unethical to change a whole health care system on the basis of intuition and opinion. Shouldn't we argue for clinical experiments, using hospitals, clinics, communities, and physicians, as the experimental units? The methodology needs development but the need is clear. Moses and Mosteller [46], for example, several years ago called for a study of the reason for the large variations in death rate, from hospital to hospital, found in the Halothane Study. Investigation of this question is under way, under the sponsorship of the National Academy of Science, and financed in part by the NIH. Certainly any proposals for change in hospitals that come out of this study must be checked experimentally in a randomized "clinical trial" of hospitals before they are implemented on the thousands of hospitals in this country.

6. Summary

In summary, I would call for joint efforts of all interested statisticians to do what they can to find out what the specific methodological problems are in clinical medicine and health care systems, to encourage strongly and enthusiastically the wise use of well tested statistical procedures on these questions, and to develop and demonstrate the use of new procedures where these are needed to communicate results and measure the credibility of scientific conclusions.

REFERENCES

[1] E. H. AHRENS, JR., "Mass field trials of the diet-heart question (their significance, timeliness, feasibility and applicability)," *American Heart Association Monograph Number 28*, 1969.

[2] C. ANELLO, "FDA principles on clinical investigations," *FDA Papers*, Vol. 4, No. 5 (1970), pp. 14–15.

[3] F. J. ANSCOMBE, "Sequential medical trials," *J. Amer. Statist. Assoc.*, Vol. 58 (1963), pp. 365–383.

[4] P. ARMITAGE, *Sequential Medical Trials*, Oxford, Blackwell Scientific Publications, 1960.

[5] J. BERKSON and R. P. GAGE, "Calculation of survival rates for cancer," *Proceedings of the Staff Meetings of the Mayo Clinic*, Vol. 25 (1950), pp. 270–286.

[6] J. W. BOAG, "Maximum likelihood estimates of the proportion of patients cured by cancer therapy," *J. Roy. Statist. Soc. Ser. B*, Vol. 11 (1949), pp. 15–53.

[7] N. BRESLOW, "A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship," *Biometrika*, Vol. 57 (1970), pp. 579–594.

[8] I. BROSS, "Two choice selection," *J. Amer. Statist. Assoc.*, Vol. 45 (1950), pp. 530–540.

[9] ———, "Scientific strategies for deep models in the growth and spread of human cancer," Symposium on the Statistical Aspects of Protocol Design, San Juan, Puerto Rico, December, 1970. Available through the Clinical Investigations Branch, National Cancer Institute, Bethesda.

[10] B. W. BROWN, JR., "The use of controls in the clinical evaluation of cancer therapies," Symposium on the Statistical Aspects of Protocol Design, San Juan, Puerto Rico, December, 1970. Available through the Clinical Investigations Branch, National Cancer Institute, Bethesda.

[11] B. W. BROWN, JR., I. FRANTZ, H. C. McGILL, JR., G. W. McMILLAN, and J. P. STRONG, "The effect of accuracy of endpoint determination on the design of therapeutic trials with particular reference to coronary heart disease," Presented at Conference on Cardiovascular Disease Epidemiology, sponsored by the Council on Epidemiology, American Heart Association, San Diego, California, March 1 and 2, 1971.

[12] J. P. BUNKER, W. H. FORREST, JR., F. MOSTELLER, and L. D. VANDAM (editors), *The National Halothane Study*, Washington, D.C., U.S. Government Printing Office, 1969.

[13] P. L. CANNER, "Selecting one of two treatments when the responses are dichotomous," *J. Amer. Statist. Assoc.*, Vol. 65 (1970), pp. 293–306.

[14] T. C. CHALMERS, "A challenge to clinical investigators," *Gastroenterology*, Vol. 57 (1969), pp. 631–635.

[15] T. C. CHALMERS, "A clinician looks at the role of of statistics in clinical research," Symposium on the Statistical Aspects of Protocol Design, San Juan, Puerto Rico, December, 1970. Available through the Clinical Investigations Branch, National Cancer Institute, Bethesda.

[16] W. G. COCHRAN and G. M. Cox, *Experimental Designs*, New York, Wiley, 1957 (2nd ed.).

[17] T. COLTON, "A model for selecting one of two medical treatments," *J. Amer. Statist. Assoc.*, Vol. 58 (1963), pp. 388–400.

[18] J. CORNFIELD, "A Bayesian test of some classical hypotheses—with applications to sequential clinical trials," *J. Amer. Statist. Assoc.*, Vol. 61 (1966), 577–594.

[19] ———, "The Bayesian outlook and its application," *Biometrics*, Vol. 25 (1969), pp. 617–58.

[20] J. CORNFIELD, M. HALPERIN, and S. W. GREENHOUSE, "An adaptive procedure for sequential clinical trials," *J. Amer. Statist. Assoc.*, Vol. 64 (1969), pp. 759–770.

[21] CORONARY DRUG PROJECT RESEARCH GROUP, "The Coronary Drug Project: initial findings leading to modifications of its research protocol," *J. Amer. Med. Assoc.*, Vol. 214 (1970), pp. 1303–1313.

[22] D. R. Cox, *Planning of Experiments*, New York, Wiley, 1958.

[23] W. J. DIXON and F. J. MASSEY, JR., *Introduction to Statistical Analysis*, New York, McGraw-Hill, 1969 (3rd ed.).

[24] F. EDERER, "A parametric estimate of the standard error of the survival rate," *J. Amer. Statist. Assoc.*, Vol. 56 (1961), pp. 111–118.

[25] B. EFRON, "The two sample problem with censored data," *Proceedings Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, Vol. 4, 1967, pp. 831–854.

[26] ———, "Forcing a sequential experiment to be balanced," Technical Report No. 14, Department of Statistics, Stanford University, 1970.

[27] A. R. Feinstein, "Clinical biostatistics, III. The architecture of clinical research," *Clinical Pharmacology and Therapeutics*, Vol. 11 (1970), pp. 432–441 (and following notes in succeeding issues of this journal).

[28] ——, "What kind of basic science for clinical medicine?", *New England J. Med.*, Vol. 283 (1970), pp. 847–852.

[29] R. A. Fisher, *Statistical Methods and Scientific Inference*, New York, Hafner 1959 (2nd ed.).

[30] ——, *The Design of Experiments*, New York, Hafner, 1966 (8th ed.). (See pp. 17–21.)

[31] E. A. Gehan, "Estimating survival functions from the life table," *J. Chronic Diseases*, Vol. 21 (1969), pp. 629–644.

[32] J. E. Grizzle, "The two-period change-over design and its use in clinical trials," *Biometrics*, Vol. 21 (1965), pp. 467–480.

[33] M. Halperin, E. Rogot, J. Gurian, and F. Ederer, "Sample sizes for medical trials with special reference to long-term therapy," *J. Chronic Diseases*, Vol. 21 (1968), pp. 13–24.

[34] J. L. Haybittle, "A two-parameter model for the survival curve of treated cancer patients," *J. Amer. Statist. Assoc.*, Vol. 60 (1965), pp. 16–26.

[35] A. B. Hill (editor), *Controlled Clinical Trials*, Oxford, Blackwell Scientific Publications, 1960.

[36] ——, "The clinical trial," *Brit. Med. Bull.*, Vol. 7 (1951), pp. 278–282.

[37] ——, *Principles of Medical Statistics*, New York, Oxford University Press, 1971 (9th ed.).

[38] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *J. Amer. Statist. Assoc.*, Vol. 53 (1958), pp. 457–481.

[39] T. J. Kiresuk, S. Salazin, and R. Sherman, "Goal attainment scaling: progress with implementation at a metropolitan community mental health center," pre-publication copy, 1971.

[40] C. R. Klimt, G. L. Knatterud, C. L. Meinert, and T. E. Prout, "The University Group Diabetes Program: A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes," *Diabetes*, Vol. 19, Supp. 2 (1970).

[41] L. Lasagna, "The controlled clinical trial: theory and practice," *J. Chronic Diseases*, Vol. 1 (1955), pp. 353–367.

[42] A. S. Littell, "Estimation of the T-year survival rate from follow-up studies over a limited period of time," *Hum. Biol.*, Vol. 24 (1952), pp. 87–116.

[43] D. Mainland, *Elementary Medical Statistics*, Philadelphia and London, W. B. Saunders, 1963.

[44] R. J. Maurice, "A different loss function for the choice between two populations," *J. Roy. Statist. Soc. Series B*, Vol. 21 (1959), pp. 203–213.

[45] R. G. Miller, Jr., "A sequential signed-rank test," *J. Amer. Statist. Assoc.*, Vol. 65 (1970), pp. 1554–1561.

[46] L. E. Moses and F. Mosteller, "Institutional differences in postoperative death rates," *J. Amer. Med. Assoc.*, Vol. 203 (1968), pp. 492–494.

[47] M. R. Novick and J. E. Grizzle, "A Bayesian approach to the analysis of data from clinical trials," *J. Amer. Statist. Assoc.*, Vol. 60 (1965), pp. 81–96.

[48] D. B. Owen, *Handbook of Statistical Tables*, Reading, Addison-Wesley, 1962.

[49] B. S. Pasternack and H. S. Gilbert, "Planning the duration of long-term survival time studies designed for accrual by cohorts," *J. Chronic Diseases*, Vol. 24 (1971), pp. 681–700.

[50] M. A. Schneiderman, "The proper size of a clinical trial: 'Grandma's Strudel' method," *J. New Drugs*, Vol. 4 (1964), pp. 3–11.

[51] G. W. Snedecor, *Statistical Methods*, Ames, Iowa State University Press, 1937 (1st ed.).

[52] G. W. Snedecor and G. W. Cochran, *Statistical Methods*, Ames, Iowa State University Press, 1967 (6th ed.).

[53] M. SOBEL and G. H. WEISS, "Play-the-winner sampling for selecting the better of two binomial populations," *Biometrika*, Vol. 57 (1970), pp. 357–65.

[54] S. M. STIGLER, "The use of random allocation for the control of selection bias," *Biometrika*, Vol. 56 (1969), pp. 553–560.

[55] VETERANS ADMINISTRATION COOPERATIVE UROLOGICAL RESEARCH GROUP, "Carcinoma of prostate: treatment comparisons," *J. Urology*, Vol. 98 (1967), pp. 516–22.

[56] L. J. WITTS (editor), *Medical Surveys and Clinical Trials*, London, Oxford University Press, 1964 (2nd ed.).

[57] WORLD HEALTH ORGANIZATION, "International drug monitoring (the role of the hospital)," Report of a WHO Meeting, World Health Organization Technical Report Series, No. 425, 1969.

[58] ———, "Principles for pre-clinical testing of drug safety," Report of a WHO Scientific Group, World Health Organization Technical Report Series, No. 341, 1966.

[59] ———, "Principles for the clinical evaluation of drugs," Report of a WHO Scientific Group, World Health Organization Technical Report Series, No. 403, 1968.

[60] M. ZELEN, "Play the winner rule and the controlled clinical trial," *J. Amer. Statist. Assoc.*, Vol. 64 (1969), pp. 131–46.

[61] M. ZELEN and P. FEIGL, "Estimation of exponential survival probabilities with concomitant information," *Biometrics*, Vol. 21, No. 4 (1965), pp. 826–38.

[62] C. ZIPPIN and P. ARMITAGE, "Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter," *Biometrics*, Vol. 22 (1966), pp. 665–72.

[63] E. GEHAN, "A generalized Wilcoxon test for comparing arbitrarily single censored samples," *Biometrika*, Vol. 52 (1965), pp. 203–223.

[64] J. BERKSON and R. P. GAGE, "Survival curve for cancer patients following treatment," *J. Amer. Statist. Assoc.*, Vol. 47, No. 259 (1952), pp. 501–515.